1. A market data acquisition system, comprising:

a means for retrieving event and embedded content data from a plurality of set-top boxes;

a means for retrieving content attributes from a content attribute database;

a means for correlating retrieved set-top box event data with content attributes to produce data indicating which content was experienced through the plurality of set-top boxes;

a means for retrieving demographic information from a demographic information database; and

a means for correlating demographic information to data indicating which content was experienced through the plurality of set-top boxes to produce, in response to a query, data indicating content experienced by at least one demographic group.

2. The market data acquisition system of Claim 1, in which said state-change data collection means collects data from said set-top boxes without access to set-top box specific personal or demographic information, thereby providing a layer of privacy to set-top box assignees.

3. The market data acquisition system of Claim 2, in which set-top box specific demographic or other personal data may be collected when requested or with approval given by a set-top box assignee, governmental agency, or other such authority.

4. The market data acquisition system of Claim 3, in which a list of set-top box identification numbers and zip codes or other geographic identifiers corresponding to set-top box installation points is provided to the present invention for each set-top box.

5. The market data acquisition system of Claim 1, in which said content attribute database is maintained as part of the system.

6. The market data acquisition system of Claim 1, in which said content presentation system is maintained external to the system.

7. The market data acquisition system of Claim 1, in which said demographic information database is maintained as part of the system.

8. The market data acquisition system of Claim 1, in which said demographic information database is maintained externally.

9. The market data acquisition system of Claim 1, in which said queries are entered through a graphical, command-line, or natural language interface.

Claims as Amended

10. The market data acquisition system of Claim 9, in which said queries can result in generation of reports for any time segment or set of time segments with high precision.

11. The market data acquisition system of Claim 9, in which said queries result in the generation of reports generated for individual content or for a set of content.

12. The market data acquisition system of Claim 9, in which said queries result in generation of said reports for persons fitting a demographic specification, persons fitting a demographic category, or persons fitting sets of demographic specifications and demographic categories.

13. The market data acquisition system of Claim 9, in which said queries result in reports generated for specific behaviors.

14. The market data acquisition system of Claim 9, in which said queries include one or more highly-specific times, demographic specifications, viewer behaviors, and content descriptions.

15. The market data acquisition system of Claim 9, in which said results are presented in a graphical manner, such as through a pie chart or bar graph.

16. The market data acquisition system of Claim 9, in which said results are presented as a spreadsheet or other grid.

17. The market data acquisition system of Claim 9, in which said results are presented as natural language.

18. The market data acquisition system of Claim 1, in which said content information is obtained from a source external to the present invention.

19. The market data acquisition system of Claim 1, in which said content information is embedded in content as it is presented to a set-top box.

20. A method of correlating dynamic and static datasets sharing at least one common characteristic and having an assumed relationship, and using such correlations to determine rule systems between the sets, comprising the steps of:

selecting subsets of said datasets sharing a common characteristic;

expressing the assumed relationship as a mathematical assumption;

defining an error function which describes the two datasets in terms of said mathematical assumption;

performing fitting procedures to account for errors in the assumed relationship; and

performing fitting procedures which account for errors in the definition of the common subsets.

21. The method of Claim 20, in which said dynamic data corresponds to set-top box event data.

22. The method of Claim 21, in which said static data corresponds to demographic data.

23. The method of Claim 22, in which correlations are drawn between set-top box event data and demographic to determine the relationship of demographics to content viewership.

24. A method of testing assumptions pertaining to relationships between two disparate datasets sharing at least one common aspect, comprising the steps of:

entering such assumptions through a user interface;

selecting sample data from a first dataset;

determining correlations between said selected data and data stored in a second dataset; and

establishing assumption validity based on such correlations.

25. A method of determining individual characteristics by correlating dynamic and static datasets sharing at least one common characteristic and having an assumed relationship, comprising the steps of:

selecting subsets of said datasets sharing a common characteristic;

expressing the assumed relationship as a mathematical assumption;

defining an error function which describes the two datasets in terms of said mathematical assumption;

performing fitting procedures to account for errors in the assumed relationship;

storing such correlations in an individual-specific array; and

iteratively repeating this process.

26. The method of Claim 25, in which said dynamic dataset corresponds to set-top box data.

27. The method of Claim 26, in which said static dataset corresponds to demographic data.

28. The method of Claim 27, in which said individual-specific data corresponds to a set-top box identification number or other privacy-compliant identification number.

Claims as Amended

29. The method of Claim 28, in which an IDM algorithm determines said correlations.

30. A method of dynamically determining the demographic identity of an individual operating a set-top box, comprising the steps of:

monitoring set-top box events for a plurality of set-top boxes;

correlating set-top box events with demographic characteristics;

applying IDM calculation techniques to determine probabilities for demographic characteristic and set-top box event dataset correlations;

ascribing demographic characteristic probabilities to each set-top box over time based on observed set-top box events and their relationship to such IDM probabilities;

evaluating such ascribed demographic characteristic probabilities over time through statistical analysis;

fitting probabilities ascribed to demographic characteristics to statistically determine the most likely set of constant dataset possibilities for each set-top box; and,

fitting set-top box possibility sets to IDM probability sets for a set-top box event.

31. The method for determining the demographic identities of individuals in a home, business, or other location containing a set-top box according to the method of Claim 30, further comprising the steps of:

storing said demographic identities in an array over time; and

applying statistical analyses to said array to determine predominant demographic identities for a given set-top box.

32. (Cancelled)

33. (Cancelled)

34. (Cancelled)

35. (Cancelled)

36. A method of determining the effect of content attributes on content ratings, comprising the steps of:

obtaining content attributes from embedded content information or from external sources;

recording set-top box events as content is experienced;

correlating set-top box events to content attributes; and,

analyzing such correlations over time to determine the effect of content attributes on content ratings.

Claims as Amended

37. The method of Claim 36 in which said content attributes include times at which various content attributes are presented to a set-top box, thereby allowing the present invention to provide detailed correlations between such attributes and set-top box events.

38. A method of determining the effect of content attributes on content ratings for a specific demographic group, comprising the steps of:

obtaining content attributes from embedded content information or from external sources;

recording set-top box events as content is experienced;

correlating set-top box events to content attributes;

correlating set-top box events and content attributes to demographic characteristics for each set-top box; and

analyzing such correlations over time to determine the effect of content attributes on content ratings for specific demographic groups.

39. The method of Claim 38 in which said content attributes include times at which various content attributes are presented to a set-top box, thereby allowing the present invention to provide detailed correlations between set-top box events, set-top box demographics, and content attributes.

40. (Cancelled)

41. (Cancelled)

42. (Cancelled)

43. (Cancelled)

44. (Cancelled)

45. (Cancelled)

46. (Cancelled)

47. (Cancelled)

48. (Cancelled)

49. (Cancelled)

50. (Cancelled)

51. (Cancelled)

52. (Cancelled)

Claims as Amended

53. (Cancelled)

54. (Cancelled)

55. (Cancelled)

56. (Cancelled)

57. (Cancelled)

58. (Cancelled)

59. (Cancelled)

60. (Cancelled)

61. (Cancelled)

62. (Cancelled)

63. A privacy-compliant data collection and data correlation system comprising:

a means of collecting individual-specific behavior data without knowing individual-specific demographic information pertaining to the individual about whom such data is collected;

a means of accessing demographic data for the region in which the individual resides; and

a means of correlating such individual-specific data with such demographic data to determine the demographic identity of each individual about whom data is collected.

64. The privacy-compliant data collection and data correlation system of Claim 63, wherein said individual-specific behavior data collection means is comprised of a set-top box.

65. (Cancelled)

66. (Cancelled)

67. A method of reducing the effect of sampling error and sample bias on data correlations determined between a dynamic dataset and a static dataset based on assumptions about the relationships between such data, [comprising] ~~such as~~:

creating equations to express ~~such~~ [data relationship] assumptions;

determining error functions which can assist in calculating values for each unknown variable in such equations;

creating a transformable matrix based on such functions;

inverting said matrix to apply a least-squares approach fitting method to the underlying data;

normalizing the results of said least-squares fit;

calculating Pearson-r correlations for such normalized results;

calculating aspect representation indices for each subset of data within said static dataset;

determining assumption validities for assumptions used as a basis for this process; and

combining said correlations, said aspect representation indices, and said assumption validities to create a set of data correlations and corresponding confidence intervals.

68. The method of Claim 67 in which said dynamic dataset represents set-top box event data.

69. The method of Claim 68 in which said static dataset represents demographic information.

70. The method of Claim 69 in which the assumption used to relate said set-top box event data with said demographic information is the demographic assumption.

71. The method of Claim ~~20~~ [67], in which said fitting procedures include applying additional assumptions to provide missing correlations values.

72. A method of increasing correlation result dataset specificity by reducing possibilities, consisting of the steps:

calculating correlation result dataset characterization values which fall within a predetermined confidence limit using aspect representation indices, inverse demographic matrices, recombination matrices, and specification similarity matrices;

creating a matrix of such values for all demographic characterizations for each method used;

utilizing mathematical expressions of the requirement of consistency for distinct value ranges for identical characterizations in the separate matrices, reducing each range for a given characterization to the greatest possible extent within a predetermined confidence interval; thus producing one matrix with one value range for each characterization;

possibly transforming value ranges for all characterizations within said matrix to the same statistical confidence;

iteratively reducing all ranges to the greatest possible extent by utilizing both mathematical expressions of the requirement of consistency among all value ranges in said matrix as well as constraints given by actual characterization population numbers; and

adjusting the statistical confidence if necessary to allow for further value range reduction past the point of useful iteration at a previous statistical confidence.

73. The method of Claim 72 in which said dataset correlations result from correlations of set-top box event data and demographic data.

74. The method of Claim 72 in which said dataset correlations result from correlations of demographic data and sales data.

75. The method of Claim 72 in which said dataset correlations result from correlations of set-top box data and sales data.

76. A method of fitting by convergence and similarity between a static dataset and a dynamic dataset, comprising the steps of:

defining subsets of each dataset;

determining correlations between such datasets;

performing a time-based analysis of group representations and additional correlations within said correlations;

assigning weights to such representations and additional correlations; and,

applying such weights and values to determine undefined correlation dataset values.

77. The method of Claim 76 in which said dynamic dataset represents set-top box data.

78. The method of Claim 77, in which said static dataset represents demographic data.

79. The method of Claim 78 in which said unidentified correlation dataset values represent non-sampled demographic specifications.

80. A method of invalidating set-top box events, comprising the steps of:

monitoring set-top box events;

storing such events in an array;

calculating trends in such events;

invalidating set-top box events which deviate in a statistically significant manner from observed set-top box event trends, or which match previously defined invalid set-top box events;

placing such invalidated set-top box events in an array; and

calculating trends in such invalidated set-top box events such that some long-term trends may be revalidated, and to identify new set-top box event categories to be ignored.